



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Todd R. Golub, Eric S. Lander, Jill Mesirov, Donna Slonim, and  
Pablo Tamayo

Application No.: 10/074,789

Group: 1631

Filed: February 12, 2002

Examiner: Not assigned

For: METHODS FOR CLASSIFYING SAMPLES AND ASCERTAINING  
PREVIOUSLY UNKNOWN CLASSES

CERTIFICATE OF MAILING	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to Assistant Commissioner for Patents, P.O. Box 2327, Arlington, VA 22202	
on <u>05/08/02</u>	<u>Annie Demirel</u>
Date	Signature
<u>Annie Demirel</u>	
Typed or printed name of person signing certificate	

PETITION TO ACCEPT COLOR DRAWINGS

UNDER 37 C.F.R. § 1.84(a)(2) AND AMENDMENT

Assistant Commissioner for Patents

P.O. Box 2327

Arlington, VA 22202

Sir:

Pursuant to 37 C.F.R. § 1.84(a)(2), Applicants hereby petition for acceptance of a color drawing of Figure 3B, consisting of one sheet. Three (3) copies of a color drawing of Figure 3B and a Transmittal of Substitute Drawings in Reply to Notice to File Corrected Application Papers are being submitted concurrently herewith. Figure 3B illustrates the relative expression of a gene (represented by the degree of shading) and the class to which the gene pertains (represented by the blue or red colors). The color is necessary to distinguish between the classes and to identify the degree of correlation of the gene to the particular class.

In compliance with 37 C.F.R. § 1.84(a)(2), a black and white photocopy of the enclosed color drawing is also enclosed.

Please amend the specification as follows:

Page 10, line 22, immediately below the header insert --- The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with the color drawing will be provided by the Office upon request and payment of the necessary fee.---

Please charge Attorney's Deposit Account No. 08-0380 for the petition fee as set forth in 37 C.F.R. § 1.17(h) in the amount of \$130.00 and for any additional fees that may be due in this matter. A copy of this transmittal letter is enclosed for accounting purposes.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By Risa M. Tyann (41,368)  
Antoinette G. Giugliano *for*  
Registration No. 42,582  
Telephone: (978) 341-0036  
Facsimile: (978) 341-0136

Concord, MA 01742-9133

Date:

10074789-03103

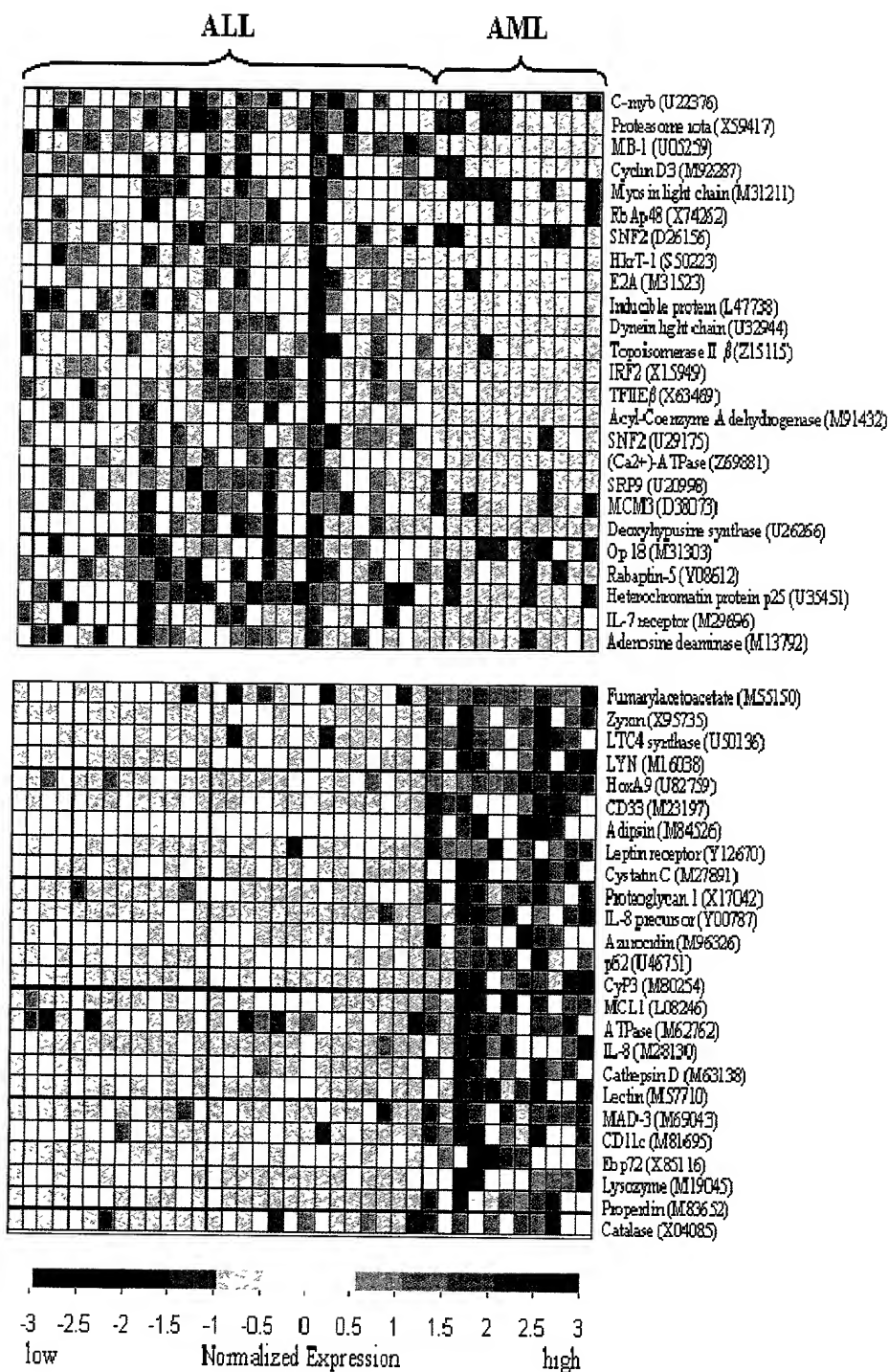


FIG. 3B

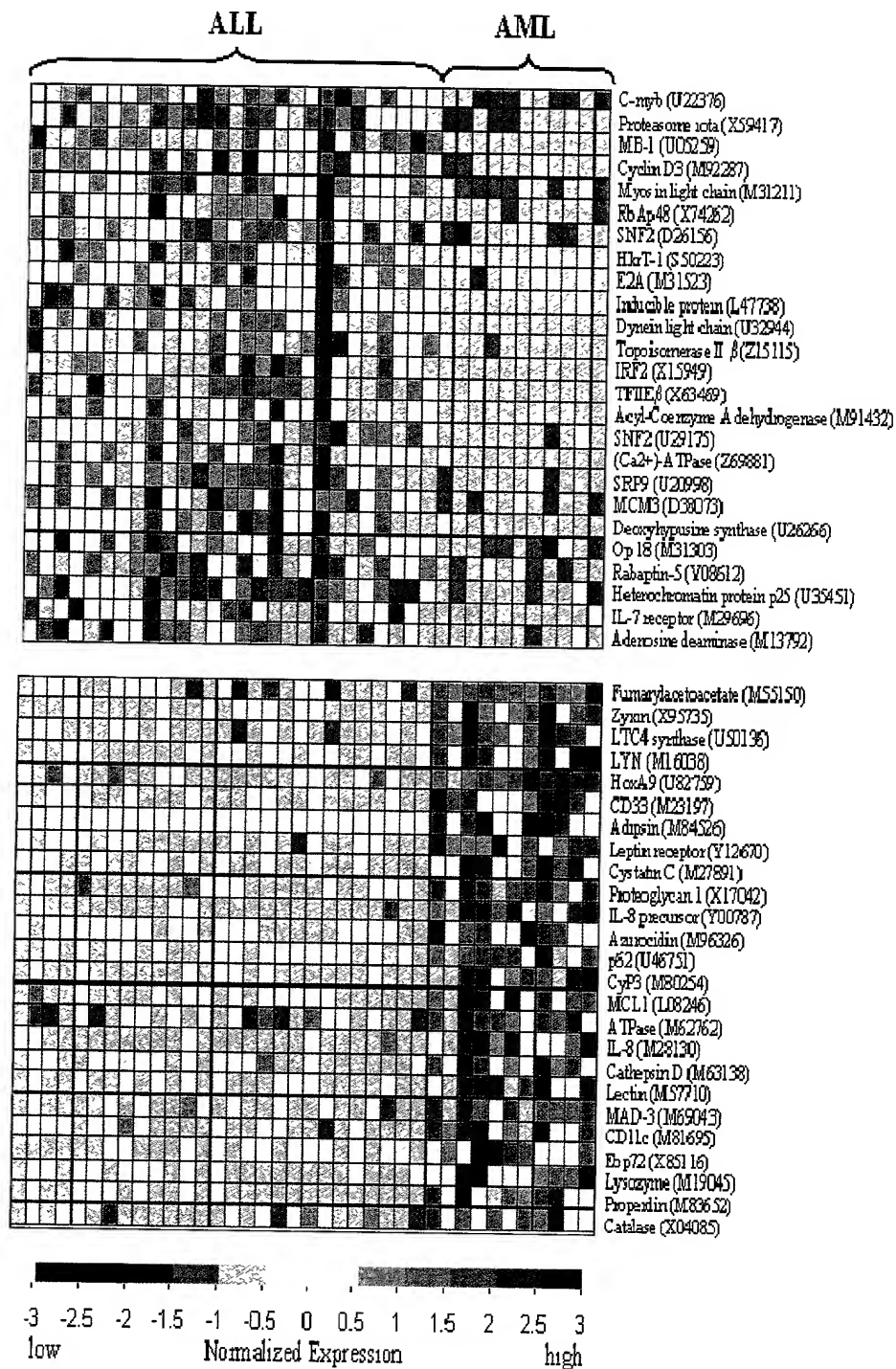


FIG. 3B

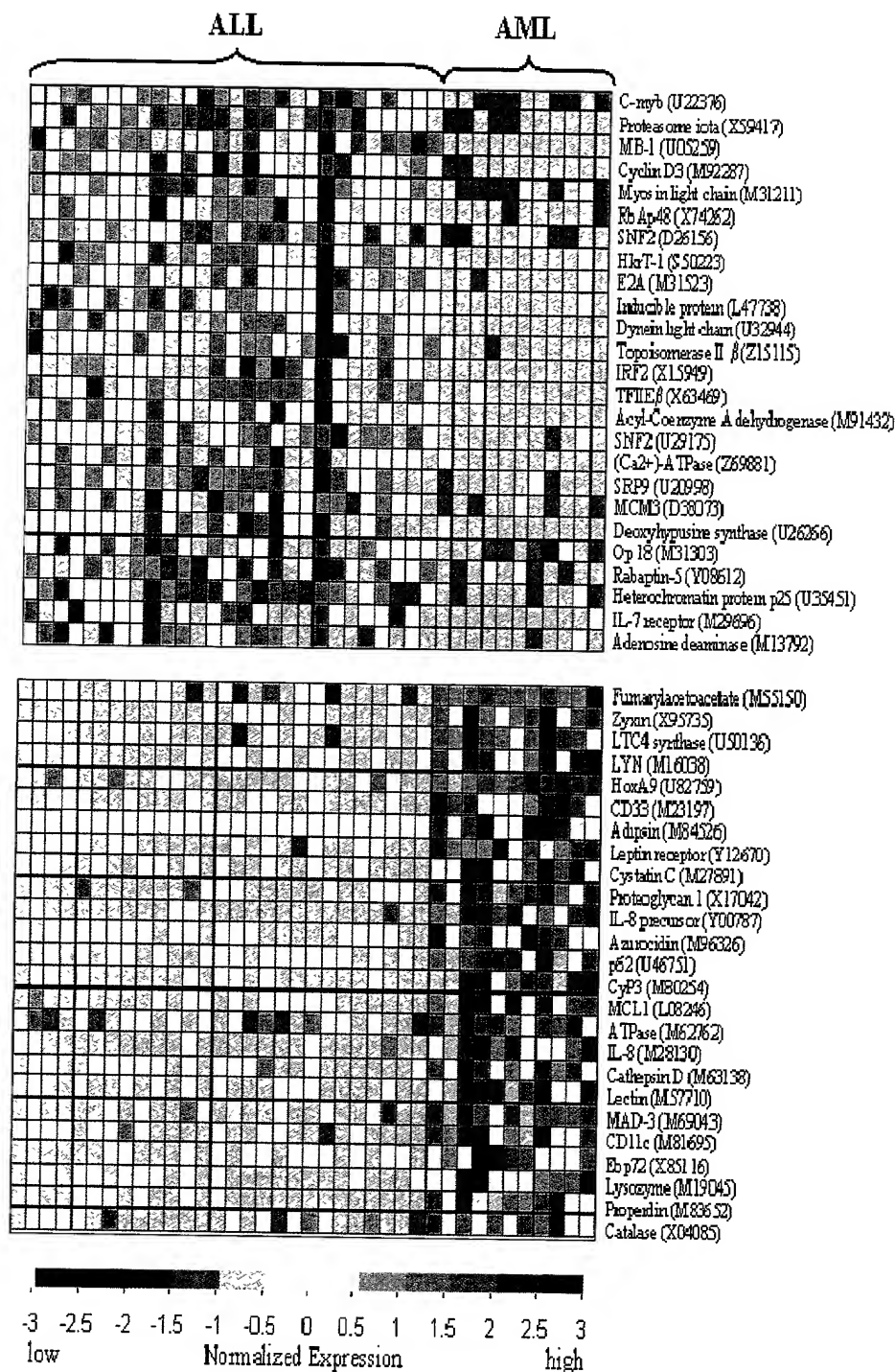


FIG. 3B

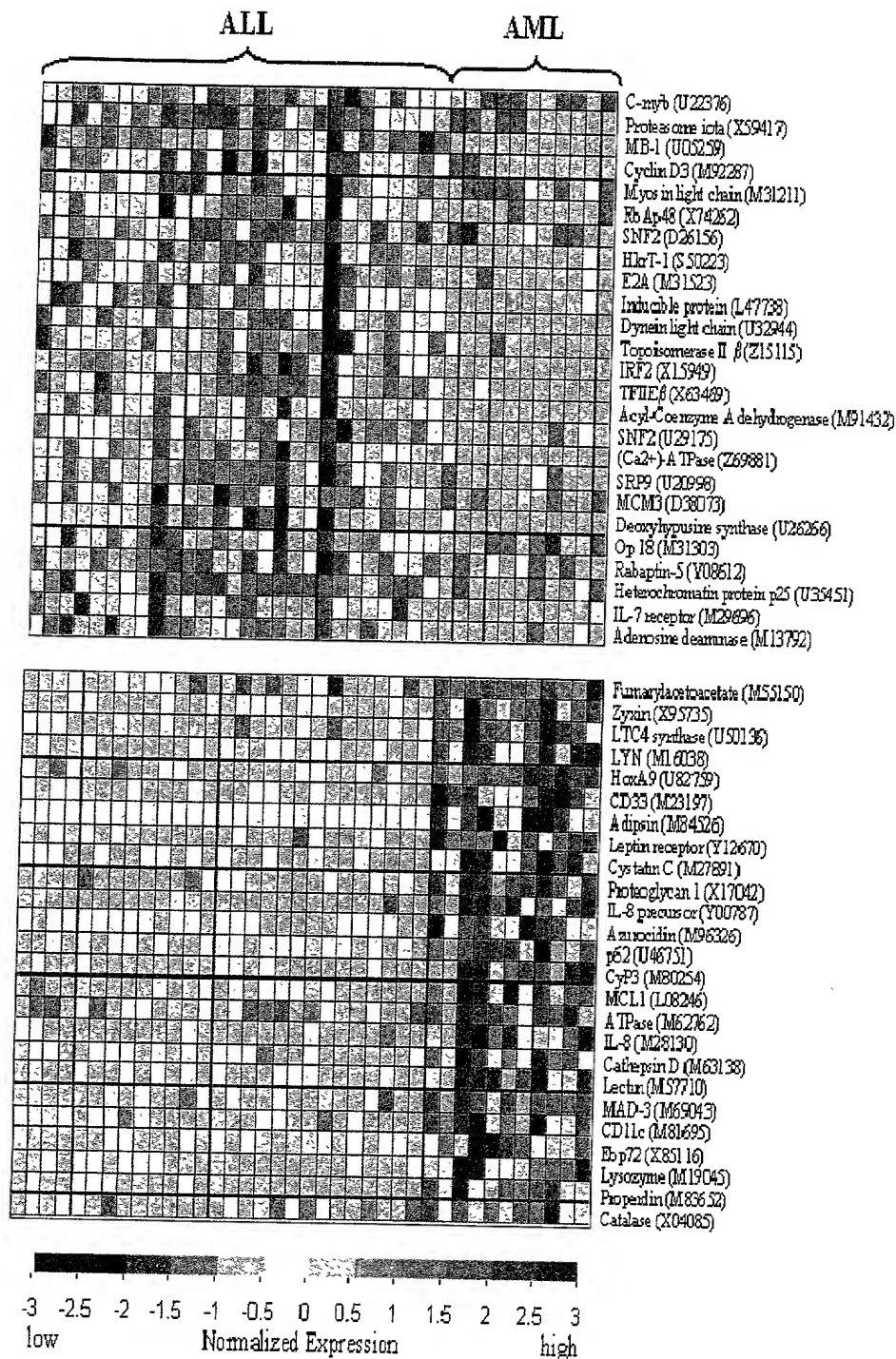
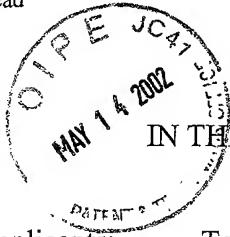


FIG. 3B



IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Todd R. Golub, Eric S. Lander, Jill Mesirov, Donna Slonim, and  
Pablo Tamayo

Application No.: 10/074,789

Group: 1631

Filed: February 12, 2002

Examiner: Not assigned

For: METHODS FOR CLASSIFYING SAMPLES AND ASCERTAINING  
PREVIOUSLY UNKNOWN CLASSES

CERTIFICATE OF MAILING	
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to Assistant Commissioner for Patents, P.O. Box 2327, Arlington, VA 22202	
on <u>05/08/02</u>	<u>Annie Demirel</u>
Date	Signature
<u>Annie Demirel</u>	
Typed or printed name of person signing certificate	

TRANSMITTAL OF SUBSTITUTE FORMAL DRAWINGS IN  
REPLY TO NOTICE TO FILE CORRECTED APPLICATION PAPERS  
AND AMENDMENT

Box Missing Parts  
Assistant Commissioner for Patents  
P.O. Box 2327  
Arlington, VA 22202

Sir:

In reply to the Notice to File Corrected Application Papers dated March 08, 2002 requesting submission of substitute drawings, please find enclosed 18 sheets of substitute drawings consisting of Figs. 1A-1C, 2A-2B, 3A-3B, 4A-4B, 5A-5D, 6, 7, 8, 9, 10, 11A-11D, and 12 for filing in the above-captioned application. Figure 10 is a black and white photograph. Figure 3B is a color drawing. A Petition to Accept Color Drawings Under 37 C.F.R. §1.84(a)(2) and a check for the petition fee as set forth in 37 C.F.R. § 1.17(h) are being filed concurrently. A copy of the Notice is attached.

- ☐ Applicant hereby petitions to extend the time to respond to the Notice to Correct Application Papers dated [ ] for [ ] month(s) from [ ] to [ ].
- ☐ The appropriate fee of \$130 is enclosed in the form of a check.
- ☒ Authorization is granted to charge Deposit Account No. 08-0380 for the petition fee due in this matter. A copy of this letter is enclosed for accounting purposes.

Please amend the application as follows:

In the Specification

Please replace the paragraph at page 11, lines 1-3 with the following paragraph:

Figures 2A-B are scatterplots showing a neighborhood analysis of genes correlating to Acute Lymphoblastic Leukemia (ALL; Figure 2A) or Acute Myeloid Leukemia (AML; Figure 2B).

Please replace the paragraph at page 11, lines 9-13 with the following paragraph:

Figures 4A-B are a set of graphs showing neighborhood analysis of genes in AML samples from patients with different clinical responses to treatment. Results are shown for 15 AML samples for which long-term clinical follow-up was available, with genes more highly expressed in the treatment failure group in Figure 4A and genes more highly expressed in the treatment success group in Figure 4B.

Please replace the paragraph at page 12, lines 10-11 with the following paragraph

Figures 11A-D are illustrations showing the assessment of statistical significance of gene-class correlations using neighborhood analysis.

Please replace the paragraph at page 41, line 20 through page 42, line 9 with the following paragraph:

The 38 acute leukemia samples were subjected to neighborhood analysis and revealed a strikingly high density of genes correlated with the AML-ALL distinction. Roughly 1100 genes were more highly correlated with the AML-ALL class distinction than would be expected by chance (Figs. 2A-B). Figures 2A-B show the number of genes within various 'neighborhoods' of the ALL/AML class distinction together with curves showing the 5% and 1% significance levels for the number of genes within corresponding neighborhoods of the randomly permuted class distinctions. Genes more highly expressed in ALL compared to AML are shown in the left panel; those more

ALL/AML distinction

highly expressed in AML compared to ALL are shown in right panel. Note the large number of genes highly correlated with the class distinction. In the left panel (higher in ALL), the number of genes with correlation  $P(g,c) > 0.30$  was 709 for the AML-ALL distinction, but had a median of 173 genes for random class distinctions. Note that  $P(g,c) = 0.30$  is the point where the observed data intersects the 1% significance level, meaning that 1% of random neighborhoods contain as many points as the observed neighborhood round the AML-ALL distinction. Similarly, in the right panel (higher in AML), 711 genes with  $P(g,c) > 0.28$  were observed, whereas a median of 136 genes is expected for random class distinctions.

Please replace the paragraph at page 46, lines 4-20 with the following paragraph:

The choice to use 50 informative genes in the predictor was somewhat arbitrary, although well within the total number of genes strongly correlated with the class distinction (Figs. 2A-B). In fact, the results proved to be quite insensitive to this choice: class predictors based on between 10 and 200 genes were tested and all were found to be 100% accurate, reflecting the strong correlation of genes with the AML-ALL distinction. Although the number of genes used had no significant effect on the outcome in this case (median PS for cross-validation ranged from 0.81 to 0.68 over a range of predictors employing 10-200 genes, all with 0% error), it may matter in other instances. One approach is to vary the number of genes used, select the number that maximizes the accuracy rate in cross-validation and then use the resulting model on the independent dataset. In any case, it is recommend that at least 10 genes be used for two reasons. Class predictors employing a small number of genes may depend too heavily on any one gene and can produce spuriously high prediction strengths (because a large 'margin of victory' can occur by chance due to statistical fluctuation resulting from a small number of genes). In general, the 1% confidence line in neighborhood analysis was also considered to be the upper bound for gene selection.

Please replace the paragraph at page 48, line 20 through page 49, line 10 with the following paragraph:

The ability to predict response to chemotherapy among the 15 adult AML patients who had been treated with an anthracycline-cytarabine regimen and for whom long-term clinical follow-up was available was explored. Treatment failure was defined as failure to achieve a complete remission following a standard induction regimen including 3 days of anthracycline and 7 days of cytarabine. Treatment successes were defined as patients in continuous complete remission for a minimum of 3 years. FAB subclass M3 patients were excluded, but samples were otherwise not selected with regard to FAB criteria. Eight patients failed to achieve remission following induction chemotherapy, while the remaining seven patients remain in remission for 46-84 months. In contrast

to the situation for the AML-ALL distinction, neighborhood analysis found no striking excess of genes correlated with response to chemotherapy (Figs 4A-B). The data fall close to the mean expected from random clusters. Nonetheless, the single most highly correlated gene, HOXA9 (arrow), is biologically related to AML. As might be expected, class predictors employing 10 to 50 genes were not highly accurate in cross-validation. For example, a 10-gene predictor yielded strong predictions ( $PS > 0.3$ ) for only 40% of the samples, and of those, 67% of the predictions were incorrect. Similarly, a 50-gene predictor yielded strong predictions for 27% of the samples, and 75% of these predictions were incorrect.

Amendments to the specification are indicated in the attached "Marked Up Version of Amendments" (pages i -iii).

#### REMARKS

This Amendment is being filed in order to make the numbering of the figures in the Specification consistent with the numbering of the figures on the formal drawings submitted concurrently. In accordance with the formal drawings, references to Figure 2 in the Specification have been changed to Figures 2A-B; references to Figure 4 in the Specification have been changed 4A-B; and reference to Figure 11 in the Specification has been changed to Figures 11A-D. No new matter is added

Acceptance of the formal drawings is respectfully requested.

Please charge any deficiency or credit any overpayment in the fees that may be due in this matter to Deposit Account No. 08-0380. A copy of this letter is enclosed for accounting purposes.

Respectfully submitted,

HAMILTON, BROOK, SMITH & REYNOLDS, P.C.

By Risa M. Treannie (41,368)  
Antoinette G. Giugliano  
Registration No.: 42,582  
Telephone: (978) 341-0036  
Facsimile: (978) 341-0136

Concord, MA 01742-9133

Date:

5/8/02

## Specification Amendments Under 37 C.F.R. § 1.121(b)(1)(iii)

Figures 2A-B [is] are [graph of] scatterplots showing a neighborhood analysis of genes correlating to Acute Lymphoblastic Leukemia (ALL; Figure 2A) or Acute Myeloid Leukemia (AML; Figure 2B).

Replace the paragraph at page 11, lines 9-13 with the below paragraph marked up by way of bracketing and underlining to show the changes relative to the previous version of the paragraph.

Figures 4A-B [is] are a set of graphs showing neighborhood analysis of genes in AML samples from patients with different clinical responses to treatment. Results are shown for 15 AML samples for which long-term clinical follow-up was available, with genes more highly expressed in the treatment failure group in [the left panel] Figure 4A and genes more highly expressed in the treatment success group in [the right panel] Figure 4B.

Replace the paragraph at page 12, lines 10-11 with the below paragraph marked up by way of bracketing and underlining to show the changes relative to the previous version of the paragraph.

Figures 11A-D [is] are [an] illustrations showing the assessment of statistical significance of gene-class correlations using neighborhood analysis.

Replace the paragraph at page 41, line 20 through page 42, line 9 with the below paragraph marked up by way of bracketing and underlining to show the changes relative to the previous version of the paragraph.

The 38 acute leukemia samples were subjected to neighborhood analysis and revealed a strikingly high density of genes correlated with the AML-ALL distinction. Roughly 1100 genes were more highly correlated with the AML-ALL class distinction than would be expected by chance (Figs. 2A-B). Figures 2A-B show[s] the number of genes within various 'neighborhoods' of the ALL/AML class distinction together with curves showing the 5% and 1% significance levels for the number of genes within corresponding neighborhoods of the randomly permuted class distinctions. Genes more highly expressed in ALL compared to AML are shown in the left panel; those more highly expressed in AML compared to ALL are shown in right panel. Note the large number of genes highly correlated with the class distinction. In the left panel (higher in ALL), the number of genes with correlation  $P(g,c) > 0.30$  was 709 for the AML-ALL distinction, but had a median of 173 genes for random class distinctions. Note that  $P(g,c) = 0.30$  is the point where the observed data intersects the 1% significance level, meaning that 1% of random neighborhoods contain as many points as the observed neighborhood round the AML-ALL distinction. Similarly, in the right panel (higher in AML), 711 genes with  $P(g,c) > 0.28$  were observed, whereas a median of 136 genes is expected for random class distinctions.

Replace the paragraph at page 46, lines 4-20 with the below paragraph marked up by way of bracketing and underlining to show the changes relative to the previous version of the paragraph.

The choice to use 50 informative genes in the predictor was somewhat arbitrary, although well within the total number of genes strongly correlated with the class distinction (Figs. 2A-B). In fact, the results proved to be quite insensitive to this choice: class predictors based on between 10 and 200 genes were tested and all were found to be 100% accurate, reflecting the strong correlation of genes with the AML-ALL distinction. Although the number of genes used had no significant effect on the outcome in this case (median PS for cross-validation ranged from 0.81 to 0.68 over a range of predictors employing 10-200 genes, all with 0% error), it may matter in other instances. One approach is to vary the number of genes used, select the number that maximizes the accuracy rate in cross-validation and then use the resulting model on the independent dataset. In any case, it is recommend that at least 10 genes be used for two reasons. Class predictors employing a small number of genes may depend too heavily on any one

gene and can produce spuriously high prediction strengths (because a large 'margin of victory' can occur by chance due to statistical fluctuation resulting from a small number of genes). In general, the 1% confidence line in neighborhood analysis was also considered to be the upper bound for gene selection.

Replace the paragraph at page 46, line 20 through page 49, line 10 with the below paragraph marked up by way of bracketing and underlining to show the changes relative to the previous version of the paragraph.

The ability to predict response to chemotherapy among the 15 adult AML patients who had been treated with an anthracycline-cytarabine regimen and for whom long-term clinical follow-up was available was explored. Treatment failure was defined as failure to achieve a complete remission following a standard induction regimen including 3 days of anthracycline and 7 days of cytarabine. Treatment successes were defined as patients in continuous complete remission for a minimum of 3 years. FAB subclass M3 patients were excluded, but samples were otherwise not selected with regard to FAB criteria. Eight patients failed to achieve remission following induction chemotherapy, while the remaining seven patients remain in remission for 46-84 months. In contrast to the situation for the AML-ALL distinction, neighborhood analysis found no striking excess of genes correlated with response to chemotherapy (Figs 4A-B). The data fall close to the mean expected from random clusters. Nonetheless, the single most highly correlated gene, HOXA9 (arrow), is biologically related to AML. As might be expected, class predictors employing 10 to 50 genes were not highly accurate in cross-validation. For example, a 10-gene predictor yielded strong predictions ( $PS > 0.3$ ) for only 40% of the samples, and of those, 67% of the predictions were incorrect. Similarly, a 50-gene predictor yielded strong predictions for 27% of the samples, and 75% of these predictions were incorrect.